

IBM Cloud Pak for Data System

With IBM Performance Server for PostgreSQL

Highlights

- A data platform with built-in governance
 - Simple data management and analysis
 - Highly available services for adaptability
 - Instant pre-assembled software provisioning with hardware
 - 100 percent compatibility with IBM PureData System for Analytics
-

Quickly scale and deploy a complete private cloud for your enterprise data and AI architecture

For years, companies have been accumulating data at an impressive pace. Some of this data is being duplicated across various clouds and repositories at astounding levels. This duplication leads to data silos, high costs, delayed projects and increased security risks.

Data users assigned to an analytics initiative spend as much as 60 percent of their overall time trying to locate the data they need. Organizations have existing processes to store data in multiple repositories, clouds and backend systems and have invested resources to move data to a central location where it can be managed, controlled and made accessible. However, most centralized big data projects still fail to deliver on the promise of easy and controlled data access. Turning that failure to success would allow data scientists to access data across their organization wherever it resides. This greatly reduces the amount of time spent searching for data.

As such, IBM developed Cloud Pak for Data System, a hyper-converged system that combines storage, computing and networking software. Built on a governed data and AI platform, Cloud Pak for Data System simplifies and unifies the management, governance and analysis of data. It allows you to provision and deploy data services flexibly and rapidly, tailored to specific needs. Users can automatically discover, add hardware nodes and add software from an app-like store to expand their system.

The Cloud Pak for Data System installation is designed to be as simple as using the software itself.

Hyper-converged technology makes it possible for users to scale and evolve their infrastructures simply and economically as application loads change. Cloud Pak for Data System includes a flexible microservices software architecture, field-programmable gate arrays (FPGAs) and FPGA-enabled network interface cards (NIC) for advanced AI acceleration.

Accelerate time to value with Cloud Pak for Data System

Customers are aware of the value and speed cloud provides to their applications and services, and want the same efficiency in their own data centers. Cloud Pak for Data System offers accelerated time to value, allowing an entire private cloud system to be stood up for a data and AI architecture in under four hours. That's why we also call it a "cloud-in-the-box."

Hyper-convergence enables a software-defined data center to provide cost-effective agility, scalability and security. Customers can quickly develop applications and analyze their data in a security-rich environment within their own firewalls and expand as they need rather than resort to expensive capital outlays. Most importantly, developers don't need to change how they deploy applications, whether they're working in the cloud or their own data center.

When your needs outgrow your existing installation, Cloud Pak for Data System simplifies expansion. Small increments of computing or storage capacity as needed, rather than spending excessively on capital. Based on application demand and delivery, budgeting can be completed in quarterly expansions on an ad hoc basis. When the hardware arrives, it's pre-loaded and configured with software and licenses. Once installed in the data center, the existing cluster will automatically discover new hardware and licenses, inherit all user configurations, and be ready to use in a few hours. The platform evolves based on each enterprise's needs and scalability, and it's all managed and highly available to help ensure maximum performance and productivity.

Benefits of bringing data & AI workloads to a "cloud-in-a-box" using Red Hat OpenShift

With a Docker registry, tools will always run on Cloud Pak for Data. It is natively built with Red Hat OpenShift Container Platform and containers migrate seamlessly to Cloud Pak for Data System.

Cloud Pak for Data supports three types of registries: Red Hat, its own, and organization-specific ones. With OpenShift, users can easily write and deploy applications knowing that they'll run on a platform optimized for Red Hat OpenShift. When choosing to deploy a private cloud on-premises, Cloud Pak for Data System provides optimized hardware to increase the container performance of the Red Hat cluster while speeding the time to value of data workloads.

IBM Performance Server for PostgreSQL

Built on Cloud Pak for Data System, IBM Performance Server for PostgreSQL (IPS) is 100 percent compatible with IBM PureData System for Analytics. Designed for deep analysis of complex and diverse data volumes scaling into the petabytes, Performance Server delivers actionable business insight with industry-leading speed and cost of ownership. Customers benefit from in-database analytics and hardware-accelerated machine learning while leveraging Cloud Pak for Data System's powerful AI capabilities such as IBM InfoSphere DataStage, IBM Cognos Analytics and IBM Watson Knowledge Studio. The combination of Cloud Pak for Data System and Performance Server allows customers to see hidden trends in customer behavior, uncover subtle changes in the market for competitive advantage and model the impact of changes made based on this intelligence.

Performance Server delivers proven performance, scalability, intelligence and simplicity for business needs. It requires minimal administration and tuning both for initial deployment as well as ongoing maintenance. IPS allows data scientists to spend a majority of their time analyzing results and providing insight rather than preparing data.

Performance Server delivers a distinct performance advantage over other analytic options. Asymmetric massively parallel processing (AMPP)[™] architecture combines Red Hat OpenShift and containers running on blade servers and NVMe disk storage with hardware-accelerated data filtering using field programmable gate arrays (FPGAs). This combination delivers fast query performance on complex analytic workloads, providing sophisticated analytics to drive business insight.

Data scientists can build their models using all enterprise data. Performance Server consolidates all your analytics into a single platform where your data resides. Linear scalability with little to no impact on performance allows data scientists and quantitative teams to operate inside the appliance without having to off-load massive data sets to separate infrastructure. Hardware acceleration allows data scientists to iterate and fine-tune analytical models faster to arrive at the best solution. Once the model is developed, it is seamlessly executed against the relevant data. Prediction and scoring can also be done where the data resides. Users get their predictive scores in near real-time, operationalizing advanced analytics by making it available throughout the enterprise.

Performance Server supports an extensive array of built-in analytical tools. IPS is delivered with a library of more than 200 pre-built, scalable, in-database analytic functions that execute analytics in parallel while abstracting away the complexity of parallel programming from developers, users and DBAs. IPS leverages native Cloud Pak for Data System support for Jupyter or Zeppelin notebooks, RStudio and a wide variety of other machine learning environments including SPSS and the Watson Catalog.

The analytics functionality extends to in-database geospatial analytics that are compatible with industry-standard ESRI GIS formats. Integration with existing geospatial analytic environments is simplified as a result. Because data lives in a variety of places and formats, IBM's Fluid Query allows users to query data in multiple data stores: third-party, open source, Hadoop, cloud and remote data sources. Users familiar with Fluid Query will find 100% compatibility in Performance Server.

Companies have accumulated and analyzed data in PureData System for Analytics for years – realizing the value of data-driven business decisions. With Cloud Pak for Data System and Performance Server, users benefit from a hyper-converged, containerized, microservices "cloud-in-a-box" architecture – bringing the scalability and economics of the cloud to their own data centers. This combination provides flexibility and ease of use with even faster and easier to manage analytics capabilities – a key component to every customer's Journey to AI.

Why IBM?

IBM's open information architecture for AI is built upon Cloud Pak for Data, which supports:

- Open source platforms and applications such as Docker and Kubernetes.
- Open source frameworks and databases such as Apache Hadoop and MongoDB.
- AI open source programming languages, deep learning frameworks and open source interfaces.
- In-database geospatial analytics combined with over 200 pre-built analytical and statistical functions.
- The only cloud-native, container-based, microservices system of insight.

Next steps

Get started by learning about Cloud Pak for Data at ibm.biz/icpdproduct, or by trying the [7-day free trial](#).

You can also [schedule a free consultation](#) for more information tailored to your business needs.

For more information

To learn more about Cloud for Data Pak System, visit ibm.biz/CloudPak4DataSystem.

© Copyright IBM Corporation 2019.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at <https://www.ibm.com/legal/us/en/copytrade.shtml>, and select third party trademarks that might be referenced in this document is available at https://www.ibm.com/legal/us/en/copytrade.shtml#section_4.

This document contains information pertaining to the following IBM products which are trademarks and/or registered trademarks of IBM Corporation:
IBM Cloud Pak for Data System



All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.